

Inequality between Member States in FP7 and Horizon 2020 – Insights from calculating Gini Coefficients

Inequality is one of the most burning economic and social issues, and the inequality between Member States in terms of Horizon 2020 funding received is equally a question of constant political interest (see also THINK Piece 1/2016). This paper presents a new approach to use standard economic tools such as Lorenz curves and the Gini coefficients for a deeper analysis of the distributional effects of FP7 and Horizon 2020. After some technical explanations and a look at the data used, the first findings from this approach are presented. In conclusion, four statements are formulated to illustrate the analytical potential of this approach.

0. Intro

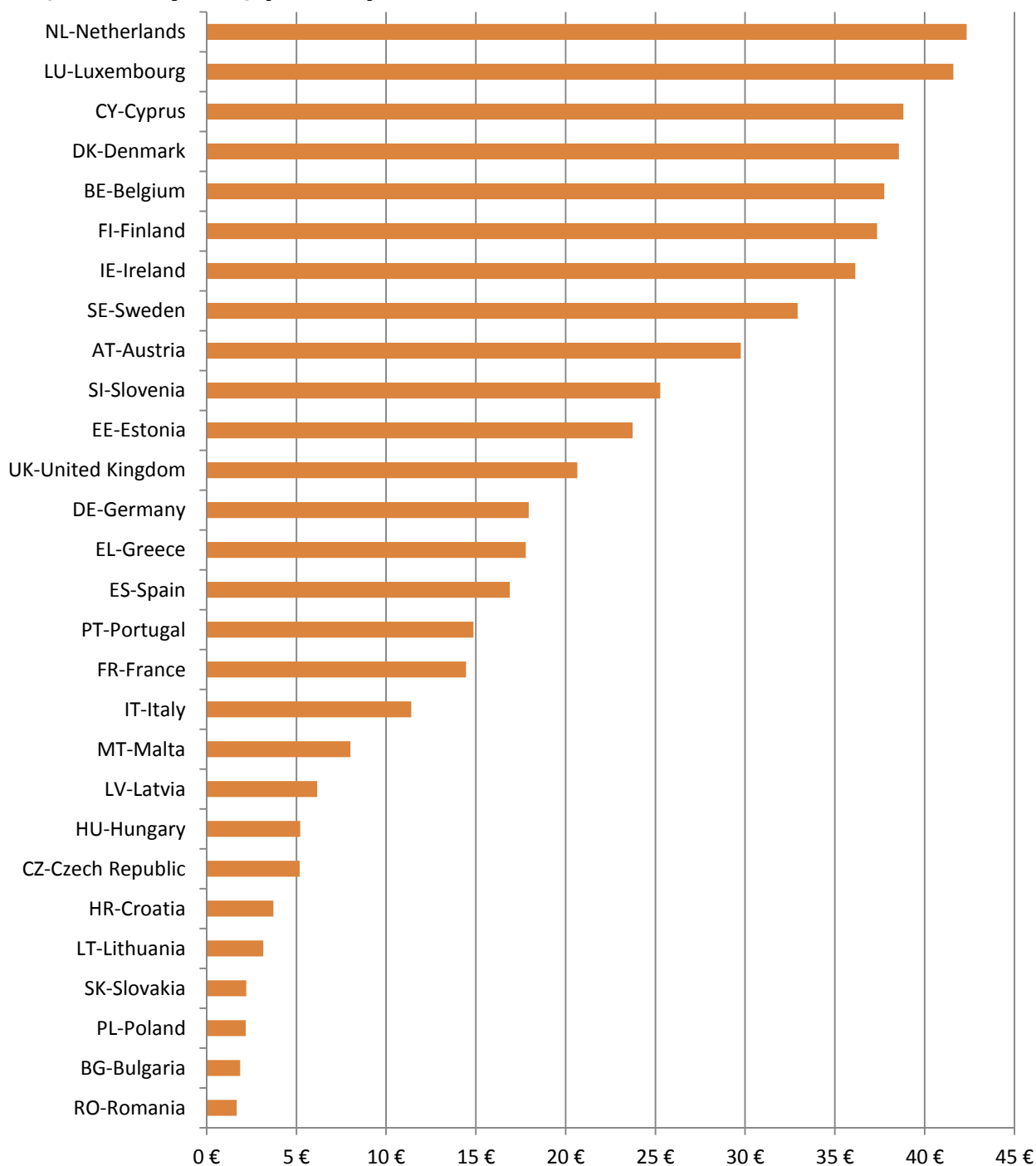
Inequality is one of the key issues in Social Sciences and also in economic theory. In very simple terms, it might be useful to distinguish two great challenges related to inequalities in our societies:

1. The first task consists of developing appropriate tools for measuring inequality. This task is based on statistical or economical concepts and dependant on adequate data, but in itself a largely non controversial issue.
2. The second task consists in interpreting these data and in developing a broader theory on the reasons and consequences of inequality. This is a largely ideological and hence very controversial question – actually one of the hot spots in the current economic debate.

The idea behind this paper is to show that tools normally used for the analysis of income (or wealth) distribution in or across societies could also be helpful when analysing the differences observed in Horizon 2020 regarding the funding received by the 28 EU Member States (see Box 1) and their contributions to the budget made (see Box 2)¹.

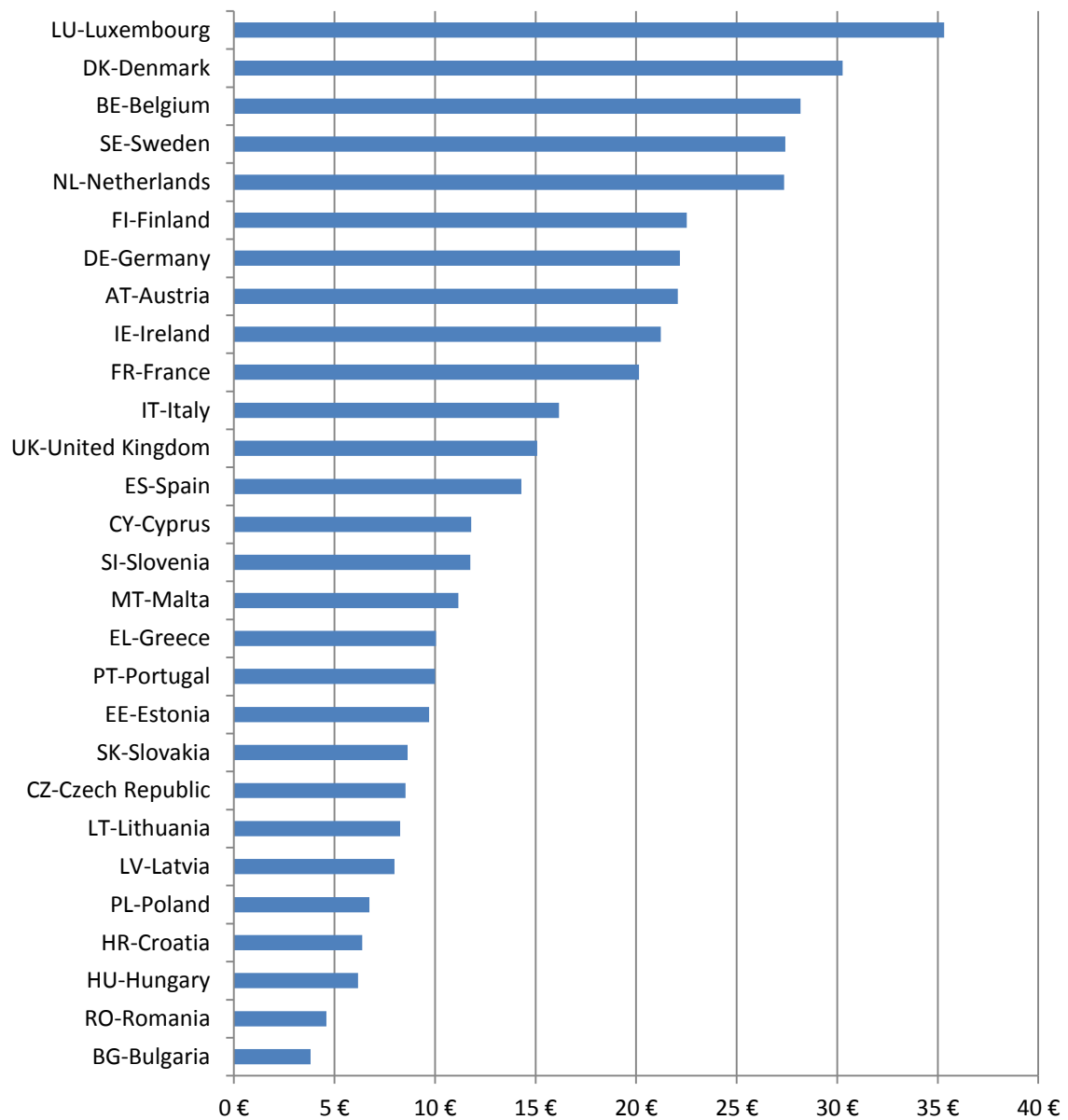
Box 1:

Estimate on income from Horizon 2020 so far ("money in") per capita



¹ The diagrams and the underlying calculations can be found in THINK Piece 1/2016

Box 2:
Estimate on spending on Horizon 2020 so far
("money out") per capita - based on 2015 budget



It seems important to state that the intention of this paper is not to enter into the normative discussion about the “right” or “optimal” level of inequality. The paper uses the term inequality to describe and analyse an uneven distribution between Member States of the funding generated from Horizon 2020 (or the contributions made to the Horizon 2020 budget). To avoid any possible misunderstandings please note that “inequality” is used in this paper as a term to describe in an absolutely neutral way a statistical phenomenon, but does by no means refer to any kind of unequal treatment or unfair results in the entire Horizon 2020 process.

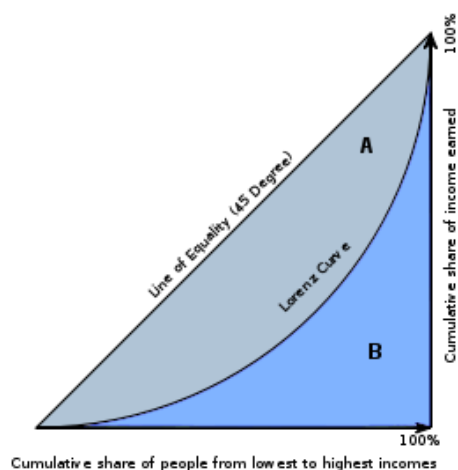
1. Tools and data

The following quote from Wikipedia² might be useful to (re-)explain the basic features:

*In economics, the **Lorenz curve** is a graphical representation of the distribution of income or of wealth. It was developed by Max O. Lorenz in 1905 for representing inequality of the wealth distribution.*

The curve is a graph showing the proportion of overall income or wealth assumed by the bottom $x\%$ of the people, although this is not rigorously true for a finite population (see below). It is often used to represent income distribution, where it shows for the bottom $x\%$ of households, what percentage ($y\%$) of the total income they have. The percentage of households is plotted on the x -axis, the percentage of income on the y -axis. It can also be used to show distribution of assets. In such use, many economists consider it to be a measure of social inequality.

The concept is useful in describing inequality among the size of individuals in ecology and in studies of biodiversity, where the cumulative proportion of species is plotted against the cumulative proportion of individuals. It is also useful in business modeling: e.g., in consumer finance, to measure the actual percentage $y\%$ of delinquencies attributable to the $x\%$ of people with worst risk scores.



Points on the Lorenz curve represent statements like "the bottom 20% of all households have 10% of the total income."

A perfectly equal income distribution would be one in which every person has the same income. In this case, the bottom $N\%$ of society would always have $N\%$ of the income. This can be depicted by the straight line $y = x$; called the "line of perfect equality."

By contrast, a perfectly unequal distribution would be one in which one person has all the income and everyone else has none. In that case, the curve would be at $y = 0\%$ for all $x < 100\%$, and $y = 100\%$ when $x = 100\%$. This curve is called the "line of perfect inequality."

*The **Gini coefficient** is the ratio of the area between the line of perfect equality and the observed Lorenz curve to the area between the line of perfect equality and the line of perfect inequality. The higher the coefficient, the more unequal the distribution is. In the diagram on the right, this is given by the ratio $A/(A+B)$, where A and B are the areas of regions as marked in the diagram.*

² https://en.wikipedia.org/wiki/Lorenz_curve

The general concept outlined above is used here to analyse the financial flows linked to the EU Framework Programmes. [THINK Piece 2/2015](#) (on FP7) and THINK Piece 1/2016 (on Horizon 2020 so far) provide for all EU Member States and for FP7 as a whole and Horizon 2020 so far data on the overall population, the funding received, both in absolute terms and per capita, as well as the contributions made to the budget, both in absolute terms and per capita³.

In addition, information provided on the distribution of funding from the European Research Council (ERC) during FP7⁴ was used to calculate the corresponding figures for this important part of the overall Framework Programme.

The data can be used to rank the entire EU population by 28 Member States as “Quantiles” from the “poorest” (country with the lowest funding or contribution per capita) to the “richest” (country with the highest funding or contribution per capita). This allows to plot Lorenz curves and to calculate Gini coefficients.

2. Results

Box 3 shows for illustration three “Lorenz curves”

- For FP7 overall (green line);
- For Horizon 2020 so far (blue line); and
- For the ERC under FP7 (Host institutions) (purple line).

The red line representing an equal distribution is shown for illustration only, notably as the calculation of the Gini coefficient depends on the size of the area between this line and the respective Lorenz curves.

Calculations for the Gini coefficient were carried out by using the online calculation tool on the website <http://www.poorcity.richcity.org/calculator/>. The raw data used and the detailed results obtained are documented in the Annex.

The Gini coefficients calculated are as follows:

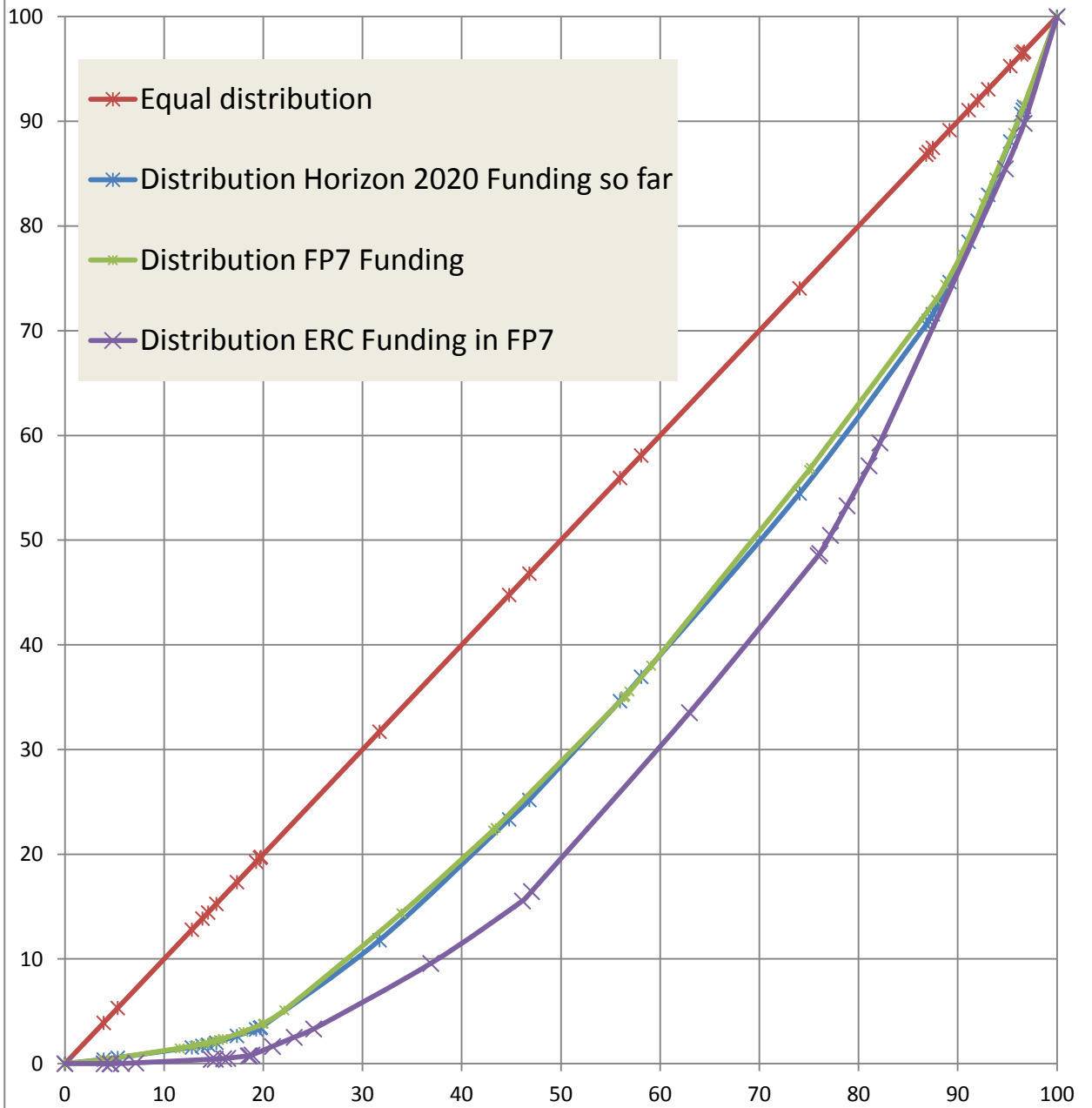
| | | | |
|---|--------------|---|--------------|
| Funding from FP7 | 31,7% | Contributions to FP7 | 23,3% |
| Funding from Horizon 2020 so far | 33,1% | Contributions to Horizon 2020 so far | 23,3% |
| Funding from ERC in FP7 | 42,3% | | |

³ The relevant data sets can be found in columns 2, 3, 5, 8 and 9 of the tables at the end of both papers. For a better understanding of the calculations, please refer in both documents also to the chapter “Data”.

⁴ ERC funding activities 2007 – 2013 – Key facts, patterns and trend, Table A8.02: Number and value of grants by current host country and funding scheme (as of 21/08/2014); available online at <https://erc.europa.eu/publication/erc-funding-activities-2007-2013>

Box 3:

"Lorenz curves" for the distribution of funding from FP7 (overall and ERC) and from Horizon 2020 so far



3. First analysis and open questions

Obviously, each of the five percentage values calculated above for the Gini coefficient of different data sets is in itself not particularly meaningful, as it is not possible to associate any tangible significance to these values. Things become, however, much more relevant when entering into direct comparisons between the different data sets under consideration. In fact, there seems to be evidence for four important statements:

- ***The contributions paid by the Member States to the budget of FP7 and Horizon 2020 so far are substantially less unequally distributed than the funding received from the Framework Programmes.***

Whereas the Gini coefficient for the contributions to FP7 and Horizon 2020 so far is 23,3%, the corresponding figure for the funding from projects is roughly some 10% higher. Since the Gini coefficient operates between two rather theoretical extreme values (0% for a totally equal distribution; 100% for a situation where one observation group alone receives (or pays) everything), a difference of 10% represents within the range of real life situations a substantive difference.

- ***The inequality of the distribution of Member States' contribution to the EU budget was basically stable from 2013 to 2015.***

The stability in the Gini coefficient at 23,3% is remarkable in the sense that recent years have seen rather important discrepancies in terms of economic development between the 28 Member States. It seems, however, that the massive differences levelled out at the end for the calculation of the contribution to the EU budget and thus to FP7 and Horizon 2020.

- ***The inequality between Member States for the funding received was in Horizon 2020 so far not lower than in FP7.***

Actually, the calculation rather suggests that there was even a marginally higher level of inequality in Horizon 2020 so far as compared to FP7 (33,1% vs. 31,7%). As this difference is relatively small and since these calculations are based on a number of assumptions and still rather preliminary data, the statement has been drafted in a rather prudent way. Nevertheless the finding is somewhat worrying, since Horizon 2020 includes a number of initiatives to foster "wider participation", which one would expect to lead to a lower level of inequality. The empirical finding not confirming this trend might mean that these measures were not yet fully implemented, or that they were insufficient to outbalance a trend towards a stronger concentration of the funding on the "stronger" Member States.

- ***The inequality between Member states for the funding received from the ERC was in FP7 substantially higher than for the overall programme.***

While it is not surprising for a research programme focused exclusively on excellence to produce a somewhat more unequal distribution of funding than FP7 as a whole, the difference in the Gini coefficients calculated here is rather substantial (42,3% vs. 31,7%). A closer look at the Lorenz curves in Box 3 suggests that the difference stems essentially from the fact that a number of Member States received virtually no funding from the ERC in FP7.

In analogy, there must be parts of FP7 (and presumably Horizon 2020) with a substantially lower Gini coefficient and a less unequal distribution of the funding across the Member States

...

4. Outlook

The new approach presented here could become a helpful tool to advance the analysis of the distributional effects of the Framework Programme. As illustrated in the previous chapter, building up proper time series for the Horizon 2020 implementation might allow identifying changes over time and possible trends. Given the political sensitivity of the issue it seems extremely important to base any future discussions on a solid empirical basis.

At the same time it seems promising to analyse not only Horizon 2020 as a whole, but to have a closer look at the major activity lines, such as the ERC. Unfortunately though, this is not yet feasible on the basis of the data published on the Open Data Portal. But again in view of forthcoming political debates on a European Innovation Council, it might be very desirable to develop a reasonable evidence base for the distributional effects of innovation programmes.

Version 1.0 – 12.03.2016 - Feedback: mail@peter-fisch.eu

Technical Annex

Table 1: Calculations for Funding from Horizon 2020 so far

| Quantile Data: People, Resources | Resources per Quantile | # | Inequalities and resource per quantile element |
|-------------------------------------|---------------------------|----|--|
| 19870647,*1,68 | 19870647.0 | 1 | |
| 7202198,*1,87 | 7202198.0 | 2 | |
| 38005614,*2,18 | 76011228.0 | 3 | |
| 5421349,*2,21 | 10842698.0 | 4 | |
| 2921262,*3,16 | 8763786.0 | 5 | |
| 4225316,*3,72 | 12675948.0 | 6 | |
| 10538275,*5,18 | 52691375.0 | 7 | |
| 9855571,*5,22 | 49277855.0 | 8 | |
| 1986096,*6,16 | 11916576.0 | 9 | |
| 429344,*8,02 | 3434752.0 | 10 | |
| 60795612,*11,39 | 668751732.0 | 11 | 508450856 quantile elements, 28 quantiles |
| 66415161,*14,45 | 929812254.0 | 12 | |
| 10374822,*14,87 | 145247508.0 | 13 | Mean: 15.754 |
| 46449565,*16,90 | 743193040.0 | 14 | Median: 0.8% 15.635 (#13/28, #14/28) |
| 10858018,*17,78 | 184586306.0 | 15 | |
| 81197537,*17,94 | 1380358129.0 | 16 | |
| 64875165,*20,66 | 1297503300.0 | 17 | Inequality Welfare |
| 1313271,*23,73 | 30205233.0 | 18 | 1-e^-TheilT: 18.7% 19.389 (1/Welfare) |
| 2062874,*25,27 | 51571850.0 | 19 | 1-e^-TheilL: 26.3% 11.617 |
| 8576261,*29,75 | 248711569.0 | 20 | 1-e^-TheilS: 22.6% 12.195 |
| 9747355,*32,92 | 311915360.0 | 21 | Gini: 33.1% 10.537 |
| 4628949,*36,14 | 166642164.0 | 22 | Plato: 35.0% 10.237 Pareto: 675/325 |
| 5471753,*37,36 | 202454861.0 | 23 | 100%-SOEP: 49.8% 7.915 |
| 11258434,*37,75 | 416562058.0 | 24 | Hoover: 21.9% 12.305 |
| 5659715,*38,56 | 215069170.0 | 25 | Theil-T Redundancy: 0.208 |
| 847008,*38,81 | 32186304.0 | 26 | Theil-L Redundancy: 0.305 |
| 562958,*41,59 | 23081278.0 | 27 | Symmetric Redundancy: 0.256 |
| 16900726,*42,35 | 709830492.0 | 28 | Inequality Issuization: +0.037 |

Table 2: Calculations for Funding from FP7

| Quantile Data: People, Resources | Resources per Quantile | # | Inequalities and resource per quantile element |
|-------------------------------------|---------------------------|----|--|
| 20020074,*7,43 | 140140518.0 | 1 | |
| 38533299,*10,37 | 385332990.0 | 2 | |
| 7284552,*13,07 | 94699176.0 | 3 | |
| 5410836,*13,36 | 70340868.0 | 4 | |
| 4262140,*17,41 | 72456380.0 | 5 | |
| 2971905,*18,54 | 53494290.0 | 6 | |
| 2023825,*20,11 | 40476500.0 | 7 | |
| 10516125,*23,71 | 241870875.0 | 8 | |
| 9908798,*28,15 | 277446344.0 | 9 | |
| 421364,*44,14 | 18540016.0 | 10 | |
| 10487289,*44,90 | 461440716.0 | 11 | 505674965 quantile elements, 28 quantiles |
| 59685227,*57,92 | 3402057939.0 | 12 | |
| 46727890,*63,09 | 2943857070.0 | 13 | Mean: 73.132 |
| 1320174,*68,32 | 89771832.0 | 14 | Median: 4.3% 69.961 (#14/28, #15/28) |
| 65578819,*70,96 | 4590517330.0 | 15 | |
| 537039,*74,11 | 39740886.0 | 16 | Inequality Welfare |
| 2058821,*79,80 | 162646859.0 | 17 | 1-e^-TheilT: 17.1% 88.200 (1/Welfare) |
| 11062508,*83,53 | 918188164.0 | 18 | 1-e^-TheilL: 22.9% 56.360 |
| 80523746,*86,53 | 6925042156.0 | 19 | 1-e^-TheilS: 20.1% 58.460 |
| 865878,*91,12 | 78794898.0 | 20 | Gini: 31.7% 49.961 |
| 63905297,*93,65 | 5943192621.0 | 21 | Plato: 32.8% 49.119 Pareto: 664/336 |
| 4591087,*116,09 | 532566092.0 | 22 | 100%-SOEP: 48.1% 37.940 |
| 8451860,*131,91 | 1107193660.0 | 23 | Hoover: 21.6% 57.368 |
| 11161642,*161,83 | 1797024362.0 | 24 | |
| 5426674,*165,50 | 895401210.0 | 25 | Theil-T Redundancy: 0.187 |
| 9555893,*166,91 | 1586278238.0 | 26 | Theil-L Redundancy: 0.261 |
| 5602628,*174,60 | 974857272.0 | 27 | Symmetric Redundancy: 0.224 |
| 16779575,*187,88 | 3137780525.0 | 28 | Inequality Issuization: +0.008 |

Table 3: Calculation for Contributions to Horizon 2020 so far

| Quantile Data: People, Resources | Resources per Quantile | # | Inequalities and resource per quantile element |
|-------------------------------------|---------------------------|----|---|
| 7202198,*3,82 | 21606594.0 | 1 | 508450856 quantile elements, 28 quantiles Mean: 16.090 Median: 3.8% 15.484 (#17/28, #18/28) Inequality Welfare 1-e^-TheilT: 9.1% 17.700 (1/Welfare) 1-e^-TheilL: 11.1% 14.299 1-e^-TheilS: 10.1% 14.461 Gini: 23.3% 12.345 Plato: 22.9% 12.407 Pareto: 614/386 100%-SOEP: 37.8% 10.015 Hoover: 16.6% 13.416 Theil-T Redundancy: 0.095 Theil-L Redundancy: 0.118 Symmetric Redundancy: 0.107 Inequality Issuization: -0.060 |
| 19870647,*4,60 | 79482588.0 | 2 | |
| 9855571,*6,18 | 59133426.0 | 3 | |
| 4225316,*6,38 | 25351896.0 | 4 | |
| 38005614,*6,73 | 228033684.0 | 5 | |
| 1986096,*7,98 | 13902672.0 | 6 | |
| 2921262,*8,27 | 23370096.0 | 7 | |
| 10538275,*8,53 | 84306200.0 | 8 | |
| 5421349,*8,64 | 43370792.0 | 9 | |
| 1313271,*9,71 | 11819439.0 | 10 | |
| 10374822,*10,00 | 103748220.0 | 11 | |
| 10858018,*10,05 | 108580180.0 | 12 | |
| 429344,*11,16 | 4722784.0 | 13 | |
| 2062874,*11,76 | 22691614.0 | 14 | |
| 847008,*11,80 | 9317088.0 | 15 | |
| 46449565,*14,29 | 650293910.0 | 16 | |
| 64875165,*15,08 | 973127475.0 | 17 | |
| 60795612,*16,16 | 972729792.0 | 18 | |
| 66415161,*20,14 | 1328303220.0 | 19 | |
| 4628949,*21,23 | 97207929.0 | 20 | |
| 8576261,*22,08 | 188677742.0 | 21 | |
| 81197537,*22,18 | 1786345814.0 | 22 | |
| 5471753,*22,51 | 120378566.0 | 23 | |
| 16900726,*27,36 | 456319602.0 | 24 | |
| 9747355,*27,42 | 263178585.0 | 25 | |
| 11258434,*28,18 | 315236152.0 | 26 | |
| 5659715,*30,26 | 169791450.0 | 27 | |
| 562958,*35,31 | 19703530.0 | 28 | |

Table 4: Calculations for Contributions to FP7

| Quantile Data: People, Resources | Resources per Quantile | # | Inequalities and resource per quantile element |
|-------------------------------------|---------------------------|----|--|
| 7284552,*17,19 | 123837384.0 | 1 | 505674965 quantile elements, 28 quantiles Mean: 73.275 Median: 3.5% 70.731 (#17/28, #18/28) Inequality Welfare 1-e^-TheilT: 8.9% 80.445 (1/Welfare) 1-e^-TheilL: 10.6% 65.541 1-e^-TheilS: 9.7% 66.140 Gini: 23.3% 56.198 Plato: 22.4% 56.833 Pareto: 612/388 100%-SOEP: 37.8% 45.576 Hoover: 17.1% 60.763 Theil-T Redundancy: 0.093 Theil-L Redundancy: 0.112 Symmetric Redundancy: 0.102 Inequality Issuization: -0.068 |
| 20020074,*20,64 | 400401480.0 | 2 | |
| 9908798,*28,27 | 277446344.0 | 3 | |
| 4262140,*29,80 | 123602060.0 | 4 | |
| 38533299,*29,84 | 1117465671.0 | 5 | |
| 2023825,*34,06 | 68810050.0 | 6 | |
| 2971905,*35,73 | 104016675.0 | 7 | |
| 5410836,*39,94 | 211022604.0 | 8 | |
| 10516125,*40,24 | 420645000.0 | 9 | |
| 1320174,*41,79 | 54127134.0 | 10 | |
| 10487289,*43,91 | 450953427.0 | 11 | |
| 11062508,*44,57 | 486750352.0 | 12 | |
| 421364,*49,32 | 20646836.0 | 13 | |
| 865878,*52,65 | 45025656.0 | 14 | |
| 2058821,*54,13 | 111176334.0 | 15 | |
| 63905297,*64,01 | 4089939008.0 | 16 | |
| 46727890,*64,75 | 2990584960.0 | 17 | |
| 59685227,*76,35 | 4536077252.0 | 18 | |
| 4591087,*92,42 | 422380004.0 | 19 | |
| 65578819,*92,52 | 6033251348.0 | 20 | |
| 80523746,*98,43 | 7891327108.0 | 21 | |
| 8451860,*101,72 | 853637860.0 | 22 | |
| 5426674,*107,36 | 580654118.0 | 23 | |
| 16779575,*123,65 | 2063887725.0 | 24 | |
| 9555893,*131,06 | 1251821983.0 | 25 | |
| 11161642,*132,45 | 1473336744.0 | 26 | |
| 5602628,*135,75 | 756354780.0 | 27 | |
| 537039,*175,16 | 93981825.0 | 28 | |

Table 5: Calculations for funding from ERC in FP7

| Quantile Data: People, Resources | Resources per Quantile | # | Inequalities and resource per quantile element |
|-------------------------------------|---------------------------|----|--|
| 20020074,*0,00 | 0.0 | 1 | |
| 2971905,*0,00 | 0.0 | 2 | |
| 421364,*0,00 | 0.0 | 3 | |
| 5410836,*21 | 113627556.0 | 4 | |
| 7284552,*45 | 327804840.0 | 5 | |
| 38533299,*56 | 2157864744.0 | 6 | |
| 2023825,*67 | 135596275.0 | 7 | |
| 4262140,*76 | 323922640.0 | 8 | |
| 2058821,*97 | 199705637.0 | 9 | |
| 10516125,*137 | 1440709125.0 | 10 | |
| 537039,*250 | 134259750.0 | 11 | 505674965 quantile elements, 28 quantiles |
| 1320174,*323 | 426416202.0 | 12 | |
| 10487289,*496 | 5201695344.0 | 13 | Mean: 1257.928 |
| 11062508,*504 | 5575504032.0 | 14 | Median: -6.8% 1344.067 (#18/28, #19/28) |
| 9908798,*510 | 5053486980.0 | 15 | |
| 59685227,*667 | 39810046409.0 | 16 | |
| 46727890,*813 | 37989774570.0 | 17 | |
| 4591087,*1240 | 5692947880.0 | 18 | |
| 80523746,*1350 | 108707057100.0 | 19 | |
| 65578819,*1454 | 95351602826.0 | 20 | |
| 865878,*1621 | 1403588238.0 | 21 | |
| 5426674,*2022 | 10972734828.0 | 22 | |
| 8451860,*2131 | 18010913660.0 | 23 | |
| 11161642,*2173 | 24254248066.0 | 24 | |
| 5602628,*2493 | 13967351604.0 | 25 | |
| 63905297,*2605 | 166473298685.0 | 26 | |
| 9555893,*2891 | 27626086663.0 | 27 | |
| 16779575,*3859 | 64752379925.0 | 28 | |
| | | | Inequality Welfare |
| | | | 1-e^-TheilT: 27.6% 1737.954 (1/Welfare) |
| | | | 1-e^-TheilL: 100.0% 0.251 |
| | | | 1-e^-TheilS: 98.8% 15.111 |
| | | | Gini: 42.3% 725.498 |
| | | | Plato: 97.8% 27.120 Pareto: 989/ 11 |
| | | | 100%-SOEP: 59.5% 509.744 |
| | | | Hoover: 30.6% 873.134 |
| | | | Theil-T Redundancy: 0.323 |
| | | | Theil-L Redundancy: 8.520 |
| | | | Symmetric Redundancy: 4.422 |
| | | | Inequality Issuization: +4.116 |